Call Variants with SAMtools Element

Calls SNPs and INDELS with SAMtools mpileup and bcftools.

Parameters in GUI

Parameter	Description	Default value
Output variants file	The url to the file with the extracted variations.	
Reference	Specify a file with the reference sequence. The sequence will be used as reference for all datasets with NGS assemblies.	
Use reference from	Specify "File" to set a single reference sequence for all input NGS assemblies. The reference should be set in the "Reference" parameter. Specify "Input port" to be able to set different references for difference NGS assemblies. The references should be input via the "Input sequences" port (e.g. use datasets in the "Read Sequence" element).	File
Illumina-1.3 + encoding	Assume the quality is in the Illumina 1.3+ encoding (mpileup)(-6).	False
Count anomalous read pairs	Do not skip anomalous read pairs in variant calling (mpileup)(-A).	False
Disable BAQ computation	Disable probabilistic realignment for the computation of base alignment quality (BAQ). BAQ is the Phred-scaled probability of a read base being misaligned. Applying this option greatly helps to reduce false SNPs caused by misalignments (mpileup)(-B).	False
Mapping quality downgrading coefficient	Coefficient for downgrading mapping quality for reads containing excessive mismatches. Given a read with a phred-scaled probability q of being generated from the mapped position, the new mapping quality is about sqrt((INT-q)/INT)*INT. A zero value disables this functionality; if enabled, the recommended value for BWA is 50 (mpileup)(-C).	0
Max number of reads per input BAM	At a position, read maximally the number of reads per input BAM (mpileup)(-d).	250
Extended BAQ computation	Extended BAQ computation. This option helps sensitivity especially for MNPs, but may hurt specificity a little bit (mpileup)(-E).	False
BED or position list file	BED or position list file containing a list of regions or sites where pileup or BCF should be generated. (mpileup)(-l).	
Pileup region	Only generate pileup in region STR (mpileup)(-r).	
Minimum mapping quality	Minimum mapping quality for an alignment to be used (mpileup)(-q).	0
Minimum base quality	Minimum base quality for a base to be considered (mpileup)(-Q).	13
Gap extension error	Phred-scaled gap extension sequencing error probability. Reducing INT leads to longer indels (mpileup)(-e).	20
Homopolymer errors coefficient	Coefficient for modeling homopolymer errors. Given an l-long homopolymer run, the sequencing error of an indel of size s is modeled as INT*s/l. (mpileup)(-h).	100
No INDELs	Do not perform INDEL calling (mpileup)(-I).	False
Max INDEL depth	Skip INDEL calling if the average per-sample depth is above INT (mpileup)(-L).	250
Gap open error	Phred-scaled gap open sequencing error probability. Reducing INT leads to more indel calls (mpileup)(-o).	40
List of platforms for indels	Comma dilimited list of platforms (determined by @RG-PL) from which indel candidates are obtained.It is recommended to collect indel candidates from sequencing technologies that have low indel error rate such as ILLUMINA. (mpileup)(-P).	

Retain all possible alternate	Retain all possible alternate alleles at variant sites. By default, the view command discards unlikely alleles. (bcf view)(-A).	False
Indicate PL	Indicate PL is generated by r921 or before (ordering is different) (bcf view)(-F).	False
No genotype information	Suppress all individual genotype information (bcf view)(-G).	False
A/C/G/T only	Skip sites where the REF field is not A/C/G/T (bcf view)(-N).	False
List of sites	List of sites at which information are outputted (bcf view)(-l).	
QCALL likelihood	Output the QCALL likelihood format (bcf view)(-Q).	False
List of samples	List of samples to use. The first column in the input gives the sample names and the second gives the ploidy, which can only be 1 or 2. When the 2nd column is absent, the sample ploidy is assumed to be 2. In the output, the ordering of samples will be identical to the one in FILE (bcf view)(-s).	
Min samples fraction	skip loci where the fraction of samples covered by reads is below FLOAT (bcf view)(-d).	0
Per-sample genotypes	Call per-sample genotypes at variant sites. (bcf view)(-g).	True
INDEL-to-SNP Ratio	Ratio of INDEL-to-SNP mutation rate. (bcf view)(-i).	-1
Max P(ref|D)	A site is considered to be a variant if P(ref|D)	0.5
Prior allele frequency spectrum	If STR can be full, cond2, flat or the file consisting of error output from a previous variant calling run (bcf view)(-P).	full
Mutation rate	Scaled mutation rate for variant calling (bcf view)(-t).	0.001
Pair/trio calling	Enable pair/trio calling. For trio calling, option -s is usually needed to be applied to configure the trio members and their ordering. In the file supplied to the option -s, the first sample must be the child, the second the father and the third the mother. The valid values of STR are pair, trioauto, trioxd and trioxs, where pair calls differences between two input samples, and trioxd (trioxs)specifies that the input is from the X chromosome non-PAR regions and the child is a female (male) (bcf view)(-T).	
N group-1 samples	Number of group-1 samples. This option is used for dividing the samples into two groups for contrast SNP calling or association test. When this option is in use, the followingVCF INFO will be outputted: PC2, PCHI2 and QCHI2 (bcf view)(-1).	0
N permutations	Number of permutations for association test (effective only with -1) (bcf view)(-U).	0
Min P (chi^2)	Only perform permutations for P(chi^2).	0.01
Minimum RMS quality	Minimum RMS mapping quality for SNPs (varFilter) (-Q).	10
Minimum read depth	Minimum read depth (varFilter) (-d).	2
Maximum read depth	Maximum read depth (varFilter) (-D).	10000000
Alternate bases	Minimum number of alternate bases (varFilter) (-a).	2
Gap size	SNP within INT bp around a gap to be filtered (varFilter) (-w).	3
Window size	Window size for filtering adjacent gaps (varFilter) (-W).	10
Strand bias	Minimum P-value for strand bias (given PV4) (varFilter) (-1).	0.0001
BaseQ bias	Minimum P-value for baseQ bias (varFilter) (-2).	1e-100
MapQ bias	Minimum P-value for mapQ bias (varFilter) (-3).	0

End distance bias	Minimum P-value for end distance bias (varFilter) (-4).	0.0001
HWE	Minimum P-value for HWE (plus F).	0.0001
Log filtered	Print filtered variants into the log (varFilter) (-p).	False

Parameters in Workflow File

Type: call_variants

Parameter	Parameter in the GUI	Type
illumina13-encoding	Illumina-1.3+ encoding	*boolean*
use_orphan	Count anomalous read pairs	*boolean*
disable_baq	Disable BAQ computation	*boolean*
capq_thres	Mapping quality downgrading coefficient	*numeric*
max_depth	Max number of reads per input BAM	*numeric*
ext_baq	Extended BAQ computation	*boolean*
bed	BED or position list file	*string*
reg	Pileup region	*string*
min_mq	Minimum mapping quality	*numeric*
min_baseq	Minimum base quality	*numeric*
extQ	Gap extension error	*numeric*
tandemQ	Homopolymer errors coefficient	*numeric*
no_indel	No INDELs	*boolean*
max_indel_depth	Max INDEL depth	*numeric*
openQ	Gap open error	*numeric*
pl_list	List of platforms for indels	*string*
keepalt	Retain all possible alternate	*boolean*
fix_pl	Indicate PL	*boolean*
no_geno	No genotype information	*boolean*
acgt_only	A/C/G/T only	*boolean*
bcf_bed	List of sites	*string*
qcall	QCALL likelihood	*boolean*
samples	List of samples	*string*
min_smpl_frac	Min samples fraction	*numeric*
call_gt	Per-sample genotypes	*boolean*
indel_frac	INDEL-to-SNP Ratio	*numeric*
pref	Max P(ref|D)	*numeric*
ptype	Prior allele frequency spectrum	*string*
theta	Mutation rate	*numeric*
ccall	Pair/trio calling	*string*
n1	N group-1 samples	*numeric*
n_perm	N permutations	*numeric*
min_perm_p	Min P(chi^2)	*numeric*
min-qual	Minimum RMS quality	*numeric*
min-dep	Minimum read depth	*numeric*
max-dep	Maximum read depth	*numeric*
min-alt-bases	Alternate bases	*numeric*

gap-size	Gap size	*numeric*
window"	Window size	*numeric*
min-strand	Strand bias	*numeric*
min-baseQ	BaseQ bias	*string*
min-mapQ	MapQ bias	*numeric*
min-end-distance	End distance bias	*numeric*
min-hwe	HWE	*numeric*
print-filtered	Log filtered	*boolean*

Input/Output Ports

The element has 2 *input ports*:

Name in GUI: Input assembly

Name in Workflow File: in-assembly

Slots:

Slot In GUI	Slot in Workflow File	Type
Dataset name	**dataset**	*string*
Source url	**url**	*string*

Name in GUI: Input sequences

Name in Workflow File: in-sequence

Slots:

Slot In GUI	Slot in Workflow File	Type
Source url	**url**	*string*

And 1 *output port*:

Name in GUI: Output variations

Name in Workflow File: out-variations

Slots:

Slot In GUI	Slot in Workflow File	Type
Variation track	**variation-track**	*variation*